

Latente Variabelen en Structurele Vergelijkingsmodellen

Toegepaste data-analyse

Inhoudsopgave

1	Confirmatorische Factor Analyse (CFA)	4
1.1	Doelstelling	4
1.2	Het standaard CFA model	8
1.3	Unidimensioneel versus multidimensioneel meetmodel	13
1.4	Reflectieve versus Formatieve indicatoren	15
1.5	Identificatie unidimensioneel CFA model	16
1.6	Schatten van de parameters in het CFA model	22
1.7	Model fitmaten	23
1.8	Model evaluatie	27
2	Implementatie CFA in lavaan en interpretatie	29
2.1	ECR-RC voor kinderen en adolescenten	29
3	SEM in vogelvlucht	41
3.1	Inleiding	41
3.2	Predictoren van jobtevredenheid	43

3.3	Stappen om SEM uit te voeren: 2 stappen	47
3.4	Modelvergelijkingen	77

1 Confirmatorische Factor Analyse (CFA)

1.1 Doelstelling

- de primaire doelstelling van Confirmatorische Factor Analyse (CFA) is om na te gaan of een *a priori* vastgelegde factor-structuur in overeenstemming is met de data
 - de factor-structuur bepaalt het aantal latente variabelen (factoren), de relatie tussen de indicatoren en de factoren, de covarianties tussen de factoren, etc. . .
 - deze factor-structuur is gebaseerd op eerder onderzoek (EFA?) en/of theoretische argumenten
- CFA is altijd de eerste stap om later een structureel vergelijkingsmodel te fitten; zolang het ‘meetmodel’ niet adequaat is, heeft het geen zin om de structurele relaties tussen de ‘factoren’ te bestuderen

Voorbeeld

- ECR-R: Experiences in Close Relationships Scale-Revised:
 - één van de meest gebruikte maten van hechting bij volwassenen
 - 2 fundamentele dimensies in hechting:
angstige en vermijdende hechting
- specifieke maat voor kinderen en jongvolwassenen?
 - studie bij 514 kinderen tussen 10 en 14 jaar
 - originele items aangepast aan kinderen: 36 items op 7-punt likert-schaal ('1=strongly disagree' tot '7=strongly agree')

- voorbeelden items angstige hechting:
 - I'm afraid my mother will stop loving me
 - I get angry because my mother doesn't give me enough love and support
- voorbeelden items vermijdende hechting:
 - I don't like telling my mother how I feel deep down inside.
 - I prefer not to get too close to my mother.
- identieke set van items werd voorgelegd aan de kinderen over de vaders

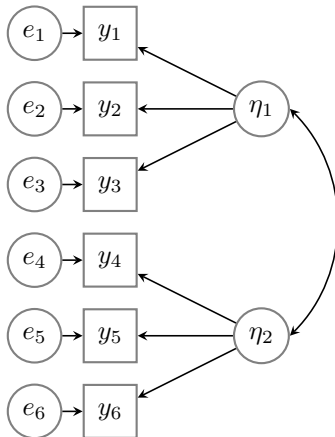
ONDERZOEKSVRAGEN:

Vinden we bij kinderen/jongvolwassenen ook een 2-delige factor-structuur?

Vinden we hier ook een goede inhoudsvaliditeit en discriminante validiteit?

- **Inhoudsvaliditeit** ('content validity') slaat op de vraag of de test- of vragenlijst items representatief zijn voor de constructen die men probeert te meten.
- **Discriminante validiteit** ('discriminant validity') is dat aspect van construct-validiteit dat betrekking heeft op de vraag of een bepaalde methode/vragenlijst in staat is verschillende begrippen van elkaar te onderscheiden (discrimineren).

1.2 Het standaard CFA model

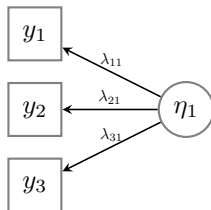


Eigenschappen 'standaard' CFA model

Een standaard CFA model heeft de volgende drie eigenschappen:

1. elke indicator y is een continue variabele die beïnvloed wordt door twee mogelijke oorzaken:
 - (a) de latente variabele (factor) waarvan de indicator een deelaspect hoort te meten
 - (b) alle overige oorzaken die een invloed kunnen hebben op y : de niet-geobserveerde error-term e
2. de error-termen zijn niet gecorreleerd met elkaar, en niet gecorreleerd met de factoren
3. alle correlaties tussen de factoren zijn niet-geanalyseerd

Elementen van het standaard CFA model: indicatoren en factoren



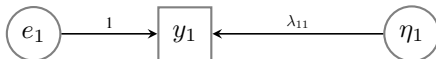
- de indicatoren beschouwen we als endogene variabelen; de factoren als exogene variabelen
- de indicatoren zijn bedoeld (deelaspecten van) de factoren te ‘meten’
 - vaak zijn het items van een vragenlijst
 - of scores op een subtesten (bvb. in een intelligentie-test)

- de factoren corresponderen met hypothetische constructen die niet (direct) observeerbaar zijn
 - vaak gaan we de factoren benoemen naar het hypothetisch construct (bvb. ‘sociale steun’ of ‘borderline personality’). Let wel: dit betekent niet noodzakelijk dat de factor dit construct meet (= *naming fallacy*)
- de schattingen van de ‘directe effecten’ van de factoren op de indicatoren noemt men *factorladingen*
- we kunnen die factorladingen interpreteren als regressiecoëfficiënten:

$$y_1 = \lambda_{11}\eta_1 + e_1$$

- de regressiecoëfficiënten kunnen we *standaardiseren* (door zowel de indicatoren als de factoren te standaardiseren)

Elementen van het standaard CFA model: error-termen



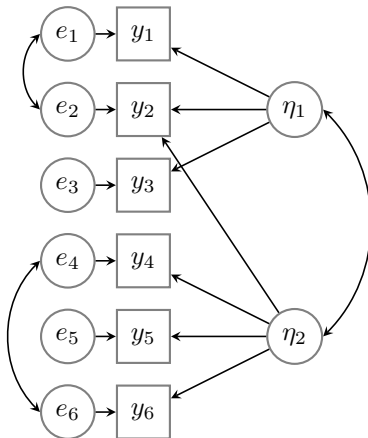
- we beschouwen de error termen als exogene variabelen (net zoals de disturbance termen in padanalyse)
- het effect van de error term op de indicator representeert het gecombineerd effect van
 - alle (weggelaten of onbekende) variabelen die een invloed hebben op de indicator (bvb. methode effecten)
 - meetfout
- de variantie van de error term noemt men de ‘unieke’ variantie van de indicator (eigen aan de indicator, versus gemeenschappelijk met de overige indicatoren)
- het direct effect van de error term op de indicator wordt gefixeerd op 1: op die manier leggen we de schaal vast van de error term

1.3 Unidimensioneel versus multidimensioneel meetmodel

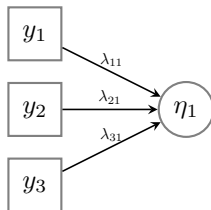
Men maakt onderscheid tussen een unidimensioneel versus een multidimensioneel meetmodel.

- unidimensioneel:
 - het standaard CFA model
 - elke indicator is gerelateerd aan slechts 1 factor
 - de error termen zijn niet gecorreleerd
- multidimensioneel:
 - het CFA model bevat mogelijks *kruisladingen*: een indicator is gerelateerd aan meer dan 1 factor
 - het CFA model bevat mogelijks *error-correlaties*: een correlatie (covariantie) tussen twee error-termen

Voorbeeld: multidimensioneel meet-model



1.4 Reflectieve versus Formatieve indicatoren



- *reflectieve* indicatoren worden beïnvloed door een factor
- *formatieve* indicatoren hebben een effect op de factor (en niet omgekeerd)
- bijvoorbeeld: de factor ‘socio-economische status’ (SES) wordt beïnvloed door geobserveerde variabelen zoals inkomen, opleiding, beroep, etc. . . .
- modellen waar dit voorkomt zijn geen CFA modellen (meer)

1.5 Identificatie unidimensioneel CFA model

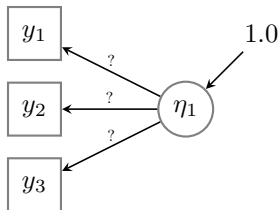
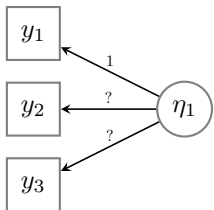
Er zijn (minstens) 3 voorwaarden opdat een CFA model geïdentificeerd zou zijn:

1. het aantal vrije parameters is kleiner dan (of gelijk aan) het aantal datapunten:
net zoals bij padanalyse is het aantal datapunten gelijk aan het aantal (niet-redundante) elementen van de variantie-covariantie matrix: $p(p + 1)/2$
2. een minimum aantal indicatoren per factor:
 - voor een CFA met (slechts) 1 factor: minstens drie indicatoren
 - voor een CFA met meerdere factoren: minstens twee indicatoren per factor
3. elke latente variabele heeft een vaste schaal
 - de error-termen: schaal wordt bepaald door de indicator (wegens direct effect gefixeerd op 1)
 - voor de factoren zijn er twee mogelijkheden: UVI of ULI

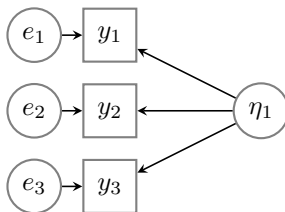
Schaal factoren: UVI of ULI

Er zijn twee courante manieren om de schaal vast te leggen van de factoren:

1. *Unit Loading Identification (ULI)*:
de niet-gestandaardiseerde regressie-coëfficiënt (factorlading) van één bepaalde indicator van deze factor wordt op 1 gefixeerd; men noemt deze indicator de *referentie* indicator
2. *Unit Variance Identification (UVI)*:
de variantie van de factor wordt op 1 gefixeerd



ULI



- aantal geobserveerde variabelen: $p = 3$
- aantal variantie/covarianties: $p(p + 1)/2 = 6$
 $\text{Var}(y_1), \text{Var}(y_2), \text{Var}(y_3)$
 $\text{Covar}(y_1, y_2), \text{Covar}(y_1, y_3), \text{Covar}(y_2, y_3)$

- we hebben

$$y_1 = 1 \cdot \eta + e_1$$

$$y_2 = \lambda_2 \eta + e_2$$

$$y_3 = \lambda_3 \eta + e_3$$

- met $\text{Covar}(\eta, e_1) = \text{Covar}(\eta, e_2) = \text{Covar}(\eta, e_3) = 0$
- veronderstel gecentreerde variabelen met gemiddelde 0
- Voor de varianties:

$$\begin{aligned}\text{Var}(y_1) &= E[y_1 y_1] \\ &= E[(\eta + e_1)(\eta + e_1)] \\ &= E(\eta\eta) + E(\eta e_1) + E(e_1 \eta) + E(e_1 e_1) \\ &= \text{Var}(\eta) + \text{Var}(e_1)\end{aligned}$$

$$\text{Var}(y_2) = \lambda_2^2 \text{Var}(\eta) + \text{Var}(e_2)$$

$$\text{Var}(y_3) = \lambda_3^2 \text{Var}(\eta) + \text{Var}(e_3)$$

- Voor de covarianties:

$$\begin{aligned}\text{Cov}(y_1, y_2) &= E[y_1 y_2] \\ &= E[(\eta + e_1)(\lambda_2 \eta + e_2)] \\ &= E(\lambda_2 \eta \eta) + E(\lambda_2 \eta e_1) + E(e_2 \eta) + E(e_1 e_2) \\ &= \lambda_2 \text{Var}(\eta)\end{aligned}$$

$$\text{Cov}(y_1, y_3) = \lambda_3 \text{Var}(\eta)$$

$$\text{Cov}(y_2, y_3) = \lambda_2 \lambda_3 \text{Var}(\eta)$$

- 6 parameters te schatten: λ_2 , λ_3 , $\text{Var}(\eta)$, $\text{Var}(e_1)$, $\text{Var}(e_2)$ en $\text{Var}(e_3)$
- aantal vrijheidsgraden=6-6=0: 'just identified' model

UVI versus ULI

- voor een standaard CFA model zou het geen verschil mogen maken
- nadeel ULI: welke indicator moeten we kiezen?
- voordeel UVI: alle factorladingen zijn vrije parameters
- nadeel UVI: enkel voor exogene factoren

1.6 Schatten van de parameters in het CFA model

- enkel via SEM software!
- standaard procedure voor continue indicatoren: Maximum Likelihood (ML) schatting vertrekkende van de variantie/covariantie matrix (niet de correlatie matrix)
- assumptie: alle indicatoren zijn van intervalniveau en bovendien multivariaat normaal verdeeld
 - indien indicatoren weliswaar continu maar scheef-verdeeld zijn, is een correctie nodig (bvb. de *Satorra-Bentler correctie*)
 - voor ‘likert scales’ wordt soms gesuggereerd om somscores te berekenen op basis van een (klein) aantal likert items die samenhangen; vervolgens hanteert men deze somscores als (continue) indicatoren; men noemt dit *item parceling*

1.7 Model fitmaten

- in principe is de interpretatie van parameters slechts zinvol indien het (pad)model de data adequaat *fit*
- SEM software rapporteert doorgaans een waaier van *fitmaten*
 - de chi-kwadraat toets
 - incrementele maten zoals de CFI, TLI
 - overige maten: RMSEA, SRMR, ...

de chi-kwadraat toets

- $\chi_M^2 = (N - 1)F_{ML}$ met N steekproefgrootte en F_{ML} geminimalizeerde fit-functie
- onder H_0 ('het model fit de data') heeft χ_M^2 een χ^2 -verdeling met het aantal vrije parameters als aantal degrees of freedom (als de steekproef voldoende groot is en uitgaande van een multivariate normaalverdeling)
- in een gesatureerd model is $\chi_M^2 = 0$: het model fit perfect de data (elke geobserveerde covariantie is gelijk aan de model-geïmpliceerde covariantie)
- als de fit van een over-geïdentificeerd model dat niet correct gespecificeerd is slechter wordt, stijgt de waarde van χ_M^2
- als χ_M^2 niet statistisch significant is, kunnen we enkel besluiten dat het model consistent is met de covariantie data, niet dat het model correct is

incrementele maten: CFI

- *CFI* Comparative Fit Index (Bentler)

- $CFI = 1 - \frac{\chi_M^2 - df_M}{\chi_B^2 - df_B}$ als $\chi_M^2 > df_M$, anders 1
- incrementele maat voor de relatieve verbetering in fit van model M versus baseline model B (model B veronderstelt geen covarianties tussen geobserveerde variabelen)
- $CFI > 0.95$ indicatie voor een goede fit

RMSEA en SRMR

- *RMSEA*: Root Mean Square Error of Approximation
 - $RMSEA = \sqrt{\frac{\chi_M^2 - df_M}{df_M(N-1)}}$ als $\chi_M^2 - df_M > 0$, anders 0
 - $RMSEA < 0.05$ indicatie voor goede fit
- *SRMR*: Standardized Root MeanSquare Residual
 - een maat voor de gemiddelde absolute correlatie residu (het verschil tussen geobserveerde en model-geïmpliceerde correlatie)
 - $SRMR < 0.08$ voor acceptabele fit

1.8 Model evaluatie

De essentie van CFA: past ('fit') het model goed met de data of niet? Drie aspecten dienen we te evalueren:

1. algemene 'goodness-of-fit' (of 'lack-of-fit') maten:

- de χ^2 statistiek is idealiter niet significant (zeldzaam met grote N)
- vuistregels: $CFI > 0.95$, $RMSEA < 0.05$, $SRMR < 0.08$
- deze vuistregels zijn vaak het onderwerp van discussie in de SEM literatuur

2. specifieke oorzaken van misfit; daarvoor bestuderen we:

- de (gestandaardiseerde) residuals (= de verschillen tussen de geobserveerde en geïmpliceerde elementen van de variantie/covariantie matrix)
- de *modification indices*

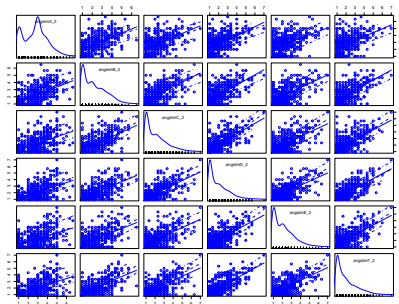
3. interpreteerbaarheid, grootte en significantie van de parameterwaarden

- ‘ongeldige’ parameterschattingen (bvb. negatieve error-varianties, negatieve factor-varianties, correlaties groter dan 1.0) zijn een signaal dat er problemen zijn (te kleine sample, slecht gespecificeerd model, ...)
- zijn de factorladingen in de ‘juiste’ richting (positief of negatief)?
- zijn alle ‘vrije parameters’ (factorladingen, factor-correlaties, error-varianties) significant? (zo niet, kunnen we ze weglaten?)
- zijn de standaardfouten van de ‘vrije parameters’ niet excessief hoog?
- is de covariantie/correlatie tussen de factoren in de juiste richting (positief of negatief)?

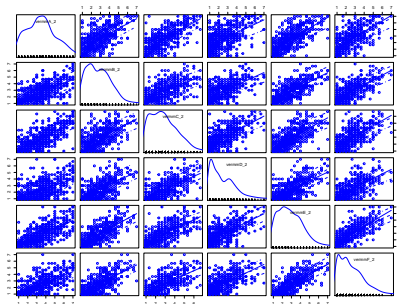
2 Implementatie CFA in lavaan en interpretatie

2.1 ECR-RC voor kinderen en adolescenten

Items die peilen naar angstige hechting:



Items die peilen naar vermijdende hechting:



- We focussen hier op de items gerelateerd aan de moeder
- *'Items assessing anxiety and avoidance were each parceled into 6 groups of 3 randomly selected items. ... Our choice for a parceling approach above an item-level CFA is informed by the current studies' sample sizes... (N=514). ... The number of parameters that has to be estimated in relation to the sample size would be out of proportion. ... A desirable goal is to have the ratio of the number of cases to the number of free parameters be 20:1 or 10:1.'*
- Satorra-Bentler correctie gebruikt.
- model-specificatie: 2 factoren met elk 6 indicatoren

ULI-approach

Code:

```
modelECR<- 'vermijding=~vermma_2+vermmB_2+vermmC_2+vermmD_2+vermmE_2+vermmF_2
          angstmA_2+angstmB_2+angstmC_2+angstmD_2+angstmE_2+angstmF_2'

fitECR1<-cfa(modelECR,data=ECR,test="satorra.bentler")

summary(fitECR1,fit.measures=T,standardized=T)
```

Output:

Estimator	ML	
Optimization method	NLMINB	
Number of free parameters	25	
	Used	Total
Number of observations	506	514

Model Test User Model:

	Standard	Robust
Test Statistic	169.384	123.621
Degrees of freedom	53	53
P-value (Chi-square)	0.000	0.000
Scaling correction factor		1.370
Satorra-Bentler correction		

Model Test Baseline Model:

Test statistic	4330.782	3228.201
Degrees of freedom	66	66

P-value	0.000	0.000
Scaling correction factor		1.342

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.973	0.978
Tucker-Lewis Index (TLI)	0.966	0.972
Robust Comparative Fit Index (CFI)		0.977
Robust Tucker-Lewis Index (TLI)		0.972

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-7921.684	-7921.684
Loglikelihood unrestricted model (H1)	-7836.992	-7836.992
Akaike (AIC)	15893.368	15893.368
Bayesian (BIC)	15999.031	15999.031
Sample-size adjusted Bayesian (BIC)	15919.678	15919.678

Root Mean Square Error of Approximation:

RMSEA	0.066	0.051
90 Percent confidence interval - lower	0.055	0.041
90 Percent confidence interval - upper	0.077	0.061
P-value RMSEA <= 0.05	0.010	0.400
Robust RMSEA		0.060
90 Percent confidence interval - lower		0.046
90 Percent confidence interval - upper		0.074

Standardized Root Mean Square Residual:

SRMR	0.041	0.041
------	-------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
vermijding = -						
vermmA_2	1.000				1.228	0.780
vermmB_2	0.861	0.043	19.912	0.000	1.057	0.811
vermmC_2	0.920	0.045	20.381	0.000	1.129	0.826
vermmD_2	0.820	0.044	18.613	0.000	1.006	0.769
vermmE_2	0.955	0.044	21.732	0.000	1.173	0.869
vermmF_2	0.900	0.042	21.568	0.000	1.105	0.864
angst = -						
angstmA_2	1.000				0.772	0.621
angstmB_2	1.188	0.082	14.446	0.000	0.917	0.788
angstmC_2	1.262	0.085	14.790	0.000	0.975	0.815
angstmD_2	1.221	0.079	15.413	0.000	0.943	0.867
angstmE_2	1.200	0.083	14.432	0.000	0.927	0.787
angstmF_2	1.211	0.080	15.199	0.000	0.935	0.849

Covariances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
vermijding --						
angst	0.577	0.065	8.867	0.000	0.609	0.609

Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.vermmA_2	0.969	0.069	14.064	0.000	0.969	0.391
.vermmB_2	0.580	0.043	13.619	0.000	0.580	0.342
.vermmC_2	0.592	0.044	13.345	0.000	0.592	0.317
.vermmD_2	0.701	0.049	14.197	0.000	0.701	0.409
.vermmE_2	0.447	0.037	12.227	0.000	0.447	0.245

.vermmF_2	0.416	0.034	12.398	0.000	0.416	0.254
.angstmA_2	0.953	0.063	15.071	0.000	0.953	0.615
.angstmB_2	0.513	0.037	13.692	0.000	0.513	0.379
.angstmC_2	0.480	0.036	13.226	0.000	0.480	0.336
.angstmD_2	0.294	0.025	11.795	0.000	0.294	0.249
.angstmE_2	0.527	0.038	13.708	0.000	0.527	0.380
.angstmF_2	0.340	0.027	12.411	0.000	0.340	0.280
vermijding	1.508	0.147	10.266	0.000	1.000	1.000
angst	0.596	0.079	7.511	0.000	1.000	1.000

- de ML-aanpak neemt observaties met ontbrekende items niet mee
- $SB-\chi^2(53) = 123.6, p < 0.001$: geen indicatie voor goede fit, maar andere fitmaten: $CFI = 0.978, RMSEA = 0.051$ OK
- ULI: 1 factorlading bij elke factor gefixeerd op 1, variantie latente vermijding en angst zijn 1.508 en 0.596
- alle gestandaardiseerde factorladingen voor angstige hechting > 0.62 en voor vermijdende hechting > 0.77
- de correlatie tussen vermijdende en angstige hechting is 0.61

UVI-approach

CODE:

```
modelECR<-'vermijding=~vermma_2+vermmB_2+vermmC_2+vermmD_2+vermmE_2+vermmF_2  
angst=~angstmA_2+angstmB_2+angstmC_2+angstmD_2+angstmE_2+angstmF_2'  
fitECR2<-cfa(modelECR,data=ECR,std.lv=T,test="satorra.bentler")  
summary(fitECR2,fit.measures=T,standardized=T)
```

OUTPUT:

Estimator	ML	
Optimization method	NLMINB	
Number of free parameters	25	
	Used	Total
Number of observations	506	514

Model Test User Model:

	Standard	Robust
Test Statistic	169.384	123.621
Degrees of freedom	53	53
P-value (Chi-square)	0.000	0.000
Scaling correction factor		1.370
Satorra-Bentler correction		

Model Test Baseline Model:

Test statistic	4330.782	3228.201
Degrees of freedom	66	66

P-value	0.000	0.000
Scaling correction factor		1.342
User Model versus Baseline Model:		
Comparative Fit Index (CFI)	0.973	0.978
Tucker-Lewis Index (TLI)	0.966	0.972
Robust Comparative Fit Index (CFI)		0.977
Robust Tucker-Lewis Index (TLI)		0.972
Loglikelihood and Information Criteria:		
Loglikelihood user model (H0)	-7921.684	-7921.684
Loglikelihood unrestricted model (H1)	-7836.992	-7836.992
Akaike (AIC)	15893.368	15893.368
Bayesian (BIC)	15999.031	15999.031
Sample-size adjusted Bayesian (BIC)	15919.678	15919.678
Root Mean Square Error of Approximation:		
RMSEA	0.066	0.051
90 Percent confidence interval - lower	0.055	0.041
90 Percent confidence interval - upper	0.077	0.061
P-value RMSEA <= 0.05	0.010	0.400
Robust RMSEA		0.060
90 Percent confidence interval - lower		0.046
90 Percent confidence interval - upper		0.074
Standardized Root Mean Square Residual:		
SRMR	0.041	0.041

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
vermijding = -						
vermmA_2	1.228	0.060	20.532	0.000	1.228	0.780
vermmB_2	1.057	0.049	21.765	0.000	1.057	0.811
vermmC_2	1.129	0.050	22.385	0.000	1.129	0.826
vermmD_2	1.006	0.050	20.094	0.000	1.006	0.769
vermmE_2	1.173	0.048	24.241	0.000	1.173	0.869
vermmF_2	1.105	0.046	24.010	0.000	1.105	0.864
angst = -						
angstmA_2	0.772	0.051	15.021	0.000	0.772	0.621
angstmB_2	0.917	0.044	20.760	0.000	0.917	0.788
angstmC_2	0.975	0.045	21.829	0.000	0.975	0.815
angstmD_2	0.943	0.039	24.041	0.000	0.943	0.867
angstmE_2	0.927	0.045	20.720	0.000	0.927	0.787
angstmF_2	0.935	0.040	23.233	0.000	0.935	0.849

Covariances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
vermijding --						
angst	0.609	0.032	19.181	0.000	0.609	0.609

Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.vermmA_2	0.969	0.069	14.064	0.000	0.969	0.391
.vermmB_2	0.580	0.043	13.619	0.000	0.580	0.342
.vermmC_2	0.592	0.044	13.345	0.000	0.592	0.317
.vermmD_2	0.701	0.049	14.197	0.000	0.701	0.409
.vermmE_2	0.447	0.037	12.227	0.000	0.447	0.245

.vermmF_2	0.416	0.034	12.398	0.000	0.416	0.254
.angstmA_2	0.953	0.063	15.071	0.000	0.953	0.615
.angstmB_2	0.513	0.037	13.692	0.000	0.513	0.379
.angstmC_2	0.480	0.036	13.226	0.000	0.480	0.336
.angstmD_2	0.294	0.025	11.795	0.000	0.294	0.249
.angstmE_2	0.527	0.038	13.708	0.000	0.527	0.380
.angstmF_2	0.340	0.027	12.411	0.000	0.340	0.280
vermijding	1.000				1.000	1.000
angst	1.000				1.000	1.000

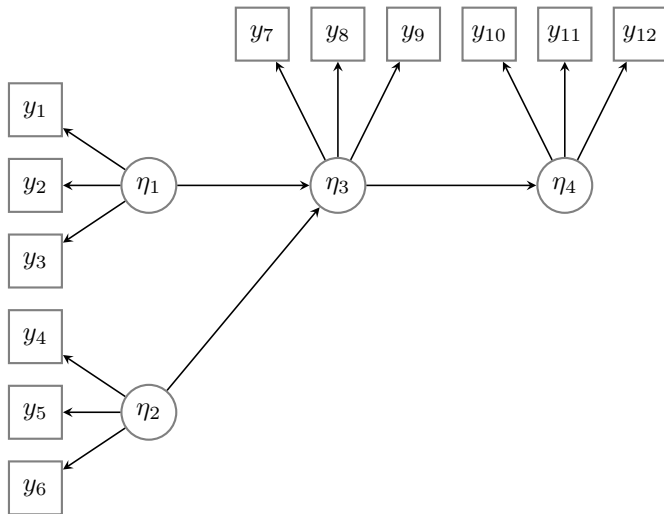
- identieke fit
- UVI: niet factorladingen gefixeerd op 1, maar varianties

3 SEM in vogelvlucht

3.1 Inleiding

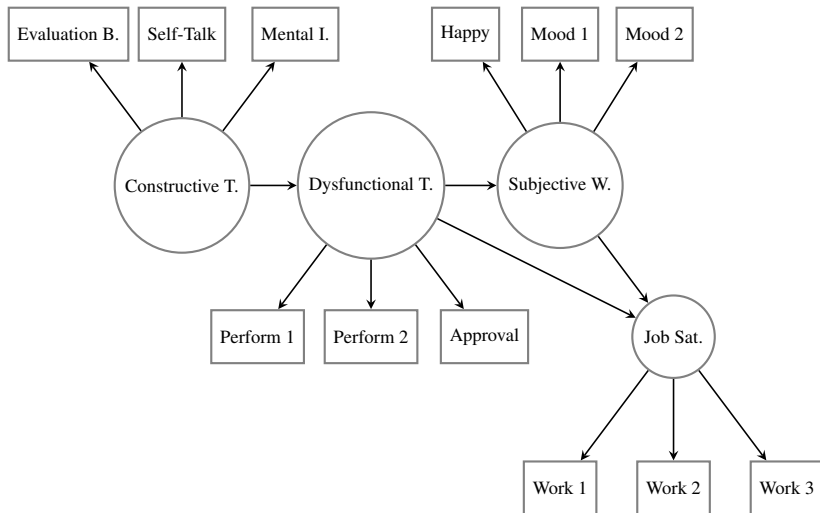
- SEM is een algemene techniek om de (lineaire) samenhang tussen variabelen te modelleren
- vertrekpunt: de samenhang tussen de geobserveerde variabelen wordt samengevat in een geobserveerde variantie/covariantie matrix (**S**)
- SEM probeert de structuur (lees: de correlatiepatronen) in deze geobserveerde variantie/covariantie matrix te beschrijven/verklaren door een model voorop te stellen met als mogelijke ingrediënten:
 - latente variabelen (cf. CFA)
 - regressies (directe en indirecte effecten) (cf. padanalyse)
 - eventueel niet-verklaarde correlaties/covarianties (dubbele pijlen)
- het model kan typisch worden weergegeven via een paddiagram

Paddiagram



3.2 Predictoren van jobtevredenheid

- onderzoek bij 263 universiteit werknemers
- 4 latente variabelen, elk gemeten door 3 indicatoren:
 - ‘constructive thinking’
 - ‘dysfunctional thinking’
 - ‘subjective well-being’
 - ‘job satisfaction’
- *“They hypothesized that constructive thinking reduces dysfunctional thinking, which leads to an enhanced sense of well-being, which in turn results in greater job satisfaction. They also predicted that dysfunctional thinking directly affects job satisfaction.”*



Geobserveerde variantie/covariantie matrix

	work1	work2	work3	happy	mood1	mood2	perform1	perform2	approval	beliefs	selftalk	imagery
work1	0.88	0.64	0.56	0.14	0.21	0.10	-0.11	-0.16	-0.18	0.05	0.06	0.11
work2	0.64	1.03	0.57	0.15	0.24	0.13	-0.11	-0.15	-0.18	0.07	0.03	0.18
work3	0.56	0.57	0.88	0.09	0.18	0.11	-0.11	-0.14	-0.13	0.06	-0.04	0.06
happy	0.14	0.15	0.09	0.32	0.21	0.07	-0.10	-0.12	-0.10	-0.01	-0.04	0.04
mood1	0.21	0.24	0.18	0.21	0.58	0.18	-0.12	-0.14	-0.14	0.08	-0.04	0.10
mood2	0.10	0.13	0.11	0.07	0.18	0.27	-0.04	-0.07	-0.08	0.05	0.00	0.02
perform1	-0.11	-0.11	-0.11	-0.10	-0.12	-0.04	0.34	0.27	0.24	-0.03	-0.03	-0.07
perform2	-0.16	-0.15	-0.14	-0.12	-0.14	-0.07	0.27	0.37	0.26	-0.05	-0.03	-0.04
approval	-0.18	-0.18	-0.13	-0.10	-0.14	-0.08	0.24	0.26	0.53	0.01	-0.01	-0.06
beliefs	0.05	0.07	0.06	-0.01	0.08	0.05	-0.03	-0.05	0.01	0.51	0.23	0.40
selftalk	0.06	0.03	-0.04	-0.04	-0.04	0.00	-0.03	-0.03	-0.01	0.23	1.26	0.43
imagery	0.11	0.18	0.06	0.04	0.10	0.02	-0.07	-0.04	-0.06	0.40	0.43	1.00

Op de diagonaal vindt men de varianties van de 12 geobserveerde indicatoren. De covarianties zijn moeilijker te interpreteren. Correlaties zijn zinnvoller:

	work1	work2	work3	happy	mood1	mood2	perform1	perform2	approval	beliefs	selftalk	imagery
work1	1.000	0.668	0.635	0.263	0.290	0.207	-0.206	-0.280	-0.258	0.080	0.061	0.113
work2	0.668	1.000	0.599	0.261	0.315	0.245	-0.182	-0.241	-0.244	0.096	0.028	0.174
work3	0.635	0.599	1.000	0.164	0.247	0.231	-0.195	-0.238	-0.185	0.094	-0.035	0.059
happy	0.263	0.261	0.164	1.000	0.486	0.251	-0.309	-0.344	-0.255	-0.017	-0.058	0.063
mood1	0.290	0.315	0.247	0.486	1.000	0.449	-0.266	-0.305	-0.255	0.151	-0.051	0.138
mood2	0.207	0.245	0.231	0.251	0.449	1.000	-0.142	-0.230	-0.215	0.141	-0.003	0.044
perform1	-0.206	-0.182	-0.195	-0.309	-0.266	-0.142	1.000	0.753	0.554	-0.074	-0.040	-0.119
perform2	-0.280	-0.241	-0.238	-0.344	-0.305	-0.230	0.753	1.000	0.587	-0.111	-0.040	-0.073
approval	-0.258	-0.244	-0.185	-0.255	-0.255	-0.215	0.554	0.587	1.000	0.016	-0.018	-0.084
beliefs	0.080	0.096	0.094	-0.017	0.151	0.141	-0.074	-0.111	0.016	1.000	0.284	0.563
selftalk	0.061	0.028	-0.035	-0.058	-0.051	-0.003	-0.040	-0.040	-0.018	0.284	1.000	0.379
imagery	0.113	0.174	0.059	0.063	0.138	0.044	-0.119	-0.073	-0.084	0.563	0.379	1.000

De fundamentele hypothese

- de fundamentele hypothese in SEM is:

$$H_0 : \Sigma = \Sigma(\theta)$$

- Σ : de populatie covariantie matrix
 - $\Sigma(\theta)$: de model-geïmpliceerde covariantie-matrix
 - θ : alle (vrije en onbekende) parameters in het model
- elk model + parameters impliceert een specifiek patroon voor de elementen van de variantie/covariantie matrix
 - we proberen een zuinig model te vinden dat goed past bij de data

3.3 Stappen om SEM uit te voeren: 2 stappen

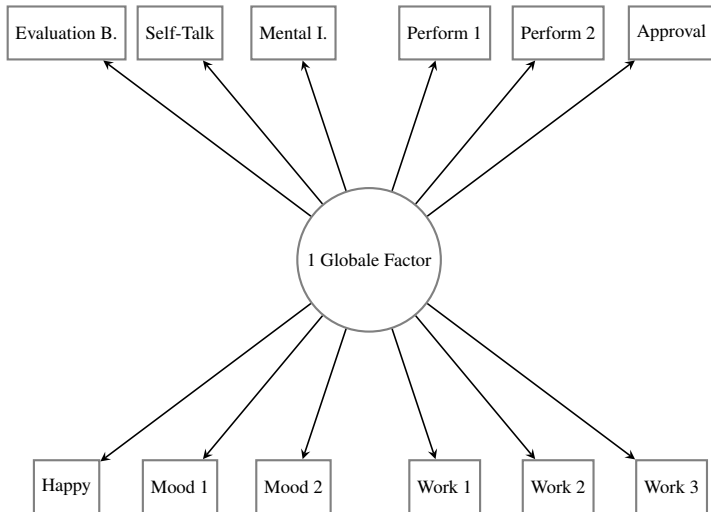
1. behandel eerst het meetmodel

- alle ‘enkelvoudige pijlen’ (regressies) tussen de latente variabelen worden vervangen door ‘dubbele pijlen’ (covarianties)
- dit is gewoon CFA
- herspecificeer het meetmodel indien nodig
- zolang de fit van het ‘meetmodel’ niet adequaat is, heeft het geen zin om de structurele component toe te voegen

2. fit het volledige model

- indien het structureel deel meer restrictief is dan het meetmodel: vergelijk het meetmodel met het volledig model (modelvergelijkingstoets)
- zijn er geen drastische verschuivingen in het meetmodel (bvb. factorladingen die plots veel groter/kleiner worden)? Zo ja, probleem!

Stap 1: Meetmodel: One-factor Model?



Implementatie met lavaan in R

```
# variantie-covariantie matrix inlezen
sds <- c(.939, 1.017, .937, .562, .760, .524, .585, .609, .731, .711, 1.124, 1.001)

COR <- '
1.000
.668 1.000
.635 .599 1.000
.263 .261 .164 1.000
.290 .315 .247 .486 1.000
.207 .245 .231 .251 .449 1.000
-.206 -.182 -.195 -.309 -.266 -.142 1.000
-.280 -.241 -.238 -.344 -.305 -.230 .753 1.000
-.258 -.244 -.185 -.255 -.255 -.215 .554 .587 1.000
.080 .096 .094 -.017 .151 .141 -.074 -.111 .016 1.000
.061 .028 -.035 -.058 -.051 -.003 -.040 -.040 -.018 .284 1.000
.113 .174 .059 .063 .138 .044 -.119 -.073 -.084 .563 .379 1.000 '

COV <- lavaan::getCov(COR, sds = sds, names = c("work1", "work2", "work3",
        "happy", "mood1", "mood2", "perform1", "perform2",
        "approval", "beliefs", "selftalk", "imagery"))

# meetmodel
modell<- 'all=~work1 + work2 + work3 + happy + mood1 + mood2 +
        perform1 + perform2 + approval + beliefs + selftalk + imagery'

# fitten model
fit1 <- cfa(modell, sample.cov=COV, sample.nobs=263)

# toon resultaten met gedetailleerde fitmaten
summary(fit1,fit.measures=T)
```

Output:

Estimator	ML
Optimization method	NLMINB
Number of free parameters	24
Number of observations	263

Model Test User Model:

Test statistic	566.797
Degrees of freedom	54
P-value (Chi-square)	0.000

Model Test Baseline Model:

Test statistic	1087.490
Degrees of freedom	66
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.498
Tucker-Lewis Index (TLI)	0.386

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-3376.334
Loglikelihood unrestricted model (H1)	-3092.936
Akaike (AIC)	6800.668
Bayesian (BIC)	6886.400
Sample-size adjusted Bayesian (BIC)	6810.309

Root Mean Square Error of Approximation:

RMSEA	0.190
90 Percent confidence interval - lower	0.176
90 Percent confidence interval - upper	0.204
P-value RMSEA <= 0.05	0.000

Standardized Root Mean Square Residual:

SRMR	0.143
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)
all =				
work1	1.000			
work2	1.026	0.202	5.072	0.000
work3	0.876	0.182	4.826	0.000
happy	0.611	0.115	5.314	0.000
mood1	0.825	0.155	5.312	0.000
mood2	0.417	0.097	4.305	0.000
perform1	-1.087	0.159	-6.826	0.000
perform2	-1.215	0.175	-6.951	0.000
approval	-1.165	0.180	-6.457	0.000
beliefs	0.230	0.116	1.972	0.049
selftalk	0.126	0.180	0.701	0.483
imagery	0.379	0.166	2.291	0.022

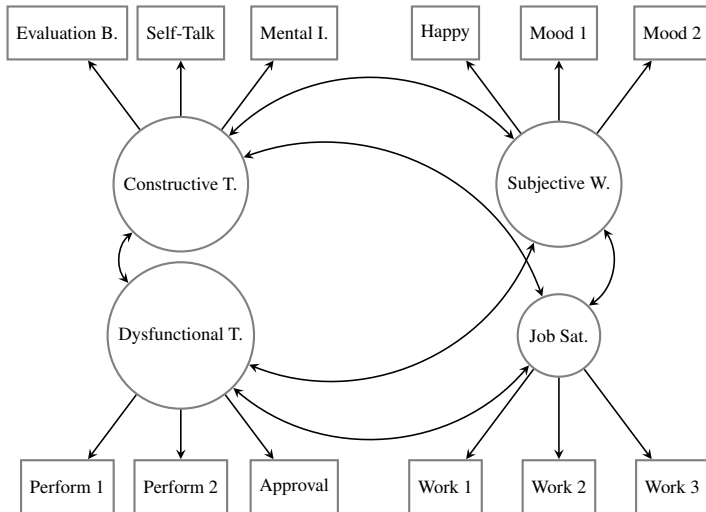
Variances:

	Estimate	Std.Err	z-value	P(> z)
.work1	0.704	0.064	10.987	0.000

.work2	0.847	0.077	11.048	0.000
.work3	0.741	0.067	11.118	0.000
.happy	0.250	0.023	10.962	0.000
.mood1	0.457	0.042	10.962	0.000
.mood2	0.243	0.022	11.227	0.000
.perform1	0.135	0.016	8.319	0.000
.perform2	0.112	0.016	6.844	0.000
.approval	0.296	0.030	9.876	0.000
.beliefs	0.494	0.043	11.432	0.000
.selftalk	1.256	0.110	11.463	0.000
.imagery	0.973	0.085	11.418	0.000
all	0.174	0.049	3.534	0.000

- aantal datapunten $p = 12 \rightarrow p \times (p + 1)/2 = 78$
- aantal parameters:
 - 11 factorladingen
 - 12 residuele varianties voor de indicatoren en 1 variantie voor de latente factor
- $df = 78 - 24 = 54$
- $\chi^2_M(54) = 566.80, p < .001$: slechte fit
- ook CFI, RMSEA, ...: indicatie voor slechte fit

Stap 1: Four-factor Model?



Implementatie met lavaan in R:

```
# meetmodel met 4 factoren
model4<- ' JobSatisfaction =~ work1 + work2 + work3
          Wellbeing      =~ happy + mood1 + mood2
          DysThinking    =~ perform1 + perform2 + approval
          ConThinking    =~ beliefs + selftalk + imagery
          ,

# fitten model
fit4 <- cfa(model4, sample.cov=COV, sample.nobs=263)

# toon resultaten met gestandaardizeerde parameterschattingen
summary(fit4,fit.measures=T,standardized=T)

# modification indices
mi4<-inspect(fit4,"mi")

subset(mi4,mi>5)
```

Output:

Estimator	ML
Optimization method	NLMINB
Number of free parameters	30
Number of observations	263

Model Test User Model:

Test statistic	62.468
----------------	--------

Degrees of freedom	48
P-value (Chi-square)	0.078

Model Test Baseline Model:

Test statistic	1087.490
Degrees of freedom	66
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.986
Tucker-Lewis Index (TLI)	0.981

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-3124.170
Loglikelihood unrestricted model (H1)	-3092.936
Akaike (AIC)	6308.340
Bayesian (BIC)	6415.505
Sample-size adjusted Bayesian (BIC)	6320.390

Root Mean Square Error of Approximation:

RMSEA	0.034
90 Percent confidence interval - lower	0.000
90 Percent confidence interval - upper	0.056
P-value RMSEA \leq 0.05	0.880

Standardized Root Mean Square Residual:

SRMR	0.040
------	-------

Parameter Estimates:

Standard errors				Standard		
Information				Expected		
Information saturated (hl) model				Structured		
Latent Variables:						
	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
JobSatisfaction =						
work1	1.000				0.786	0.839
work2	1.035	0.081	12.763	0.000	0.814	0.802
work3	0.891	0.073	12.147	0.000	0.701	0.749
Wellbeing =						
happy	1.000				0.339	0.604
mood1	1.792	0.246	7.297	0.000	0.607	0.800
mood2	0.817	0.126	6.482	0.000	0.277	0.529
DysThinking =						
perform1	1.000				0.486	0.832
perform2	1.129	0.080	14.062	0.000	0.548	0.902
approval	0.992	0.089	11.177	0.000	0.482	0.660
ConThinking =						
beliefs	1.000				0.460	0.649
selftalk	1.053	0.178	5.914	0.000	0.485	0.432
imagery	1.889	0.329	5.739	0.000	0.870	0.871
Covariances:						
	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
JobSatisfaction --						
Wellbeing	0.124	0.026	4.786	0.000	0.468	0.468
DysThinking	-0.131	0.030	-4.447	0.000	-0.344	-0.344
ConThinking	0.060	0.029	2.090	0.037	0.165	0.165
Wellbeing --						
DysThinking	-0.077	0.016	-4.817	0.000	-0.468	-0.468
ConThinking	0.025	0.013	1.896	0.058	0.162	0.162
DysThinking --						
ConThinking	-0.028	0.017	-1.640	0.101	-0.124	-0.124

Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.work1	0.261	0.042	6.229	0.000	0.261	0.297
.work2	0.368	0.050	7.399	0.000	0.368	0.357
.work3	0.384	0.044	8.684	0.000	0.384	0.439
.happy	0.200	0.022	8.993	0.000	0.200	0.635
.mood1	0.207	0.045	4.635	0.000	0.207	0.360
.mood2	0.197	0.020	9.889	0.000	0.197	0.720
.perform1	0.105	0.016	6.706	0.000	0.105	0.308
.perform2	0.069	0.017	4.013	0.000	0.069	0.186
.approval	0.300	0.030	10.178	0.000	0.300	0.564
.beliefs	0.292	0.042	6.871	0.000	0.292	0.579
.selftalk	1.023	0.097	10.507	0.000	1.023	0.813
.imagery	0.242	0.123	1.972	0.049	0.242	0.242
.JobSatisfaction	0.618	0.081	7.615	0.000	1.000	1.000
.Wellbeing	0.115	0.025	4.537	0.000	1.000	1.000
.DysThinking	0.236	0.031	7.608	0.000	1.000	1.000
.ConThinking	0.212	0.049	4.295	0.000	1.000	1.000

- model $\chi^2(48) = 62.47, p = .078$: de exact-fit hypothese niet verworpen
- ook RMSEA=0.034 en CFI=0.986 wijzen op goede fit
- het four-factor model is significant beter dan het one-factor model: $\chi^2_D(6) = 504.33, p < .001$
- kunnen we de fit nog verbeteren?

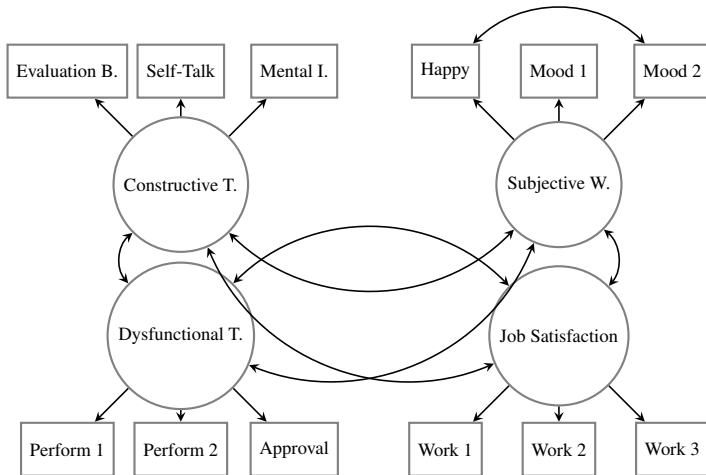
Modificatie indices (afgekort: 'mi'):

- geassocieerd met een bepaalde, gefixeerde parameter
- geeft de voorspelde daling in χ^2 -maat van het model weer als de parameter vrij gelaten wordt
- hoe groter de waarde van de mi, hoe beter de voorspelde verbetering in de fit van het model als die extra vrije parameter wordt toegevoegd
- het model herspecifiëren a.d.h.v. modificatie-indices is een voorbeeld van *data-driven* modelontwikkeling
- vaak worden mi groter dan 5 of 10 geselecteerd

	lhs op	rhs	mi	epc	sepc.lv	sepc.all	sepc.nox
56	DysThinking =	happy	6.875	-0.235	-0.114	-0.204	-0.204
102	happy --	mood2	5.380	-0.040	-0.040	-0.202	-0.202
106	happy --	beliefs	5.273	-0.040	-0.040	-0.165	-0.165
128	perform2 --	beliefs	5.161	-0.031	-0.031	-0.216	-0.216
130	perform2 --	imagery	6.185	0.047	0.047	0.363	0.363
131	approval --	beliefs	5.938	0.050	0.050	0.170	0.170

- één van de grootste χ^2 voor de covariantie tussen de indicatoren 'happy' en 'mood2' van het latente construct 'subjective well-being': dit kan wijzen op een gezamenlijke weggelaten oorzaak van deze indicatoren

Four-factor Model met bijkomende residuele correlatie



Implementatie met lavaan in R:

```
# meetmodel met 4 factoren en correlated residual

model4b<- 'JobSatisfaction =~ work1 + work2 + work3
Wellbeing      =~ happy + mood1 + mood2
DysThinking    =~ perform1 + perform2 + approval
ConThinking    =~ beliefs + selftalk + imagery
# correlated residual
happy ~~ mood2
'

# fitten model

fit4b <- cfa(model4b, sample.cov=COV, sample.nobs=263)

summary(fit4b, fit.measures=T, standardized=T)
```

Output:

Estimator	ML
Optimization method	NLMINB
Number of free parameters	31
Number of observations	263
Model Test User Model:	
Test statistic	56.662
Degrees of freedom	47
P-value (Chi-square)	0.158

Model Test Baseline Model:

Test statistic	1087.490
Degrees of freedom	66
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.991
Tucker-Lewis Index (TLI)	0.987

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-3121.267
Loglikelihood unrestricted model (H1)	-3092.936
Akaike (AIC)	6304.534
Bayesian (BIC)	6415.270
Sample-size adjusted Bayesian (BIC)	6316.985

Root Mean Square Error of Approximation:

RMSEA	0.028
90 Percent confidence interval - lower	0.000
90 Percent confidence interval - upper	0.052
P-value RMSEA <= 0.05	0.936

Standardized Root Mean Square Residual:

SRMR	0.037
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected

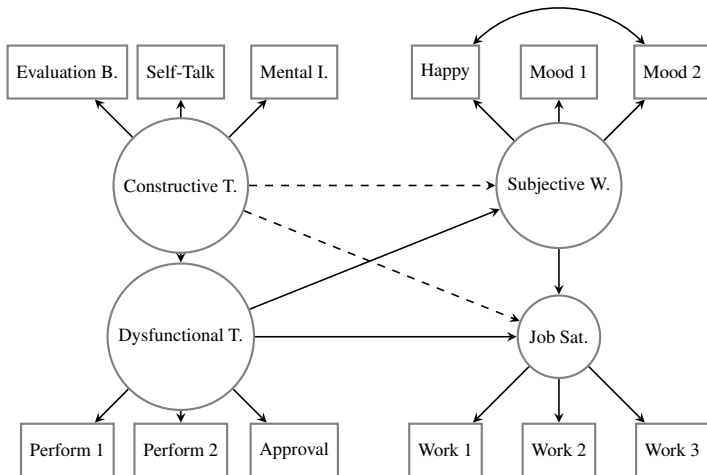
Information saturated (hl) model	Structured					
Latent Variables:	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
JobSatisfaction = "						
work1	1.000				0.786	0.839
work2	1.035	0.081	12.770	0.000	0.814	0.802
work3	0.891	0.073	12.145	0.000	0.700	0.749
Wellbeing = "						
happy	1.000				0.376	0.671
mood1	1.490	0.219	6.799	0.000	0.561	0.739
mood2	0.821	0.126	6.535	0.000	0.309	0.591
DysThinking = "						
perform1	1.000				0.485	0.830
perform2	1.133	0.080	14.105	0.000	0.549	0.904
approval	0.993	0.089	11.175	0.000	0.481	0.660
ConThinking = "						
beliefs	1.000				0.460	0.648
selftalk	1.056	0.178	5.922	0.000	0.486	0.433
imagery	1.890	0.331	5.717	0.000	0.870	0.870
Covariances:	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.happy --						
.mood2	-0.043	0.018	-2.390	0.017	-0.043	-0.243
JobSatisfaction --						
Wellbeing	0.138	0.028	5.011	0.000	0.466	0.466
DysThinking	-0.131	0.030	-4.448	0.000	-0.344	-0.344
ConThinking	0.060	0.029	2.090	0.037	0.165	0.165
Wellbeing --						
DysThinking	-0.088	0.017	-5.139	0.000	-0.480	-0.480
ConThinking	0.024	0.014	1.685	0.092	0.140	0.140
DysThinking --						
ConThinking	-0.028	0.017	-1.637	0.102	-0.124	-0.124

Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.work1	0.260	0.042	6.231	0.000	0.260	0.297
.work2	0.368	0.050	7.394	0.000	0.368	0.357
.work3	0.384	0.044	8.692	0.000	0.384	0.439
.happy	0.173	0.025	6.877	0.000	0.173	0.550
.mood1	0.261	0.044	5.970	0.000	0.261	0.453
.mood2	0.178	0.022	8.133	0.000	0.178	0.651
.perform1	0.106	0.016	6.783	0.000	0.106	0.311
.perform2	0.068	0.017	3.975	0.000	0.068	0.183
.approval	0.300	0.029	10.186	0.000	0.300	0.564
.beliefs	0.292	0.043	6.862	0.000	0.292	0.580
.selftalk	1.022	0.097	10.496	0.000	1.022	0.812
.imagery	0.242	0.123	1.965	0.049	0.242	0.242
JobSatisfaction	0.618	0.081	7.616	0.000	1.000	1.000
Wellbeing	0.142	0.031	4.643	0.000	1.000	1.000
DysThinking	0.235	0.031	7.598	0.000	1.000	1.000
ConThinking	0.212	0.049	4.285	0.000	1.000	1.000

- de exact-fit hypothese wordt niet verworpen: $\chi^2(47) = 56.66, p = .158$
- ook RMSEA, CFI, ... wijzen op goede fit
- dit four-factor model met extra correlatie heeft een significant betere fit dan het model zonder extra correlatie: $\chi^2_D(1) = 5.80, p = .016$
- nu het meetdeel is vastgelegd, kunnen we focussen op het structureel deel

SEM with Directed Arrows



Implementatie in lavaan in R:

```
semmodell<-'JobSatisfaction =~ work1 + work2 + work3
Wellbeing      =~ happy + mood1 + mood2
DysThinking    =~ perform1 + perform2 + approval
ConThinking    =~ beliefs + selftalk + imagery
# correlated residual
happy ~~ mood2
# regressies
DysThinking    ~ ConThinking
Wellbeing      ~ DysThinking + ConThinking
JobSatisfaction ~ DysThinking + ConThinking + Wellbeing
,
```

```
fitsem1 <- cfa(semmodell, sample.cov=COV, sample.nobs=263)
```

```
summary(fitsem1,fit.measures=T)
```

Output:

Estimator	ML
Optimization method	NLMINB
Number of free parameters	31

Number of observations	263
------------------------	-----

Model Test User Model:

Test statistic	56.662
Degrees of freedom	47
P-value (Chi-square)	0.158

Model Test Baseline Model:

Test statistic	1087.490
Degrees of freedom	66
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.991
Tucker-Lewis Index (TLI)	0.987

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-3121.267
Loglikelihood unrestricted model (H1)	-3092.936
Akaike (AIC)	6304.534
Bayesian (BIC)	6415.270
Sample-size adjusted Bayesian (BIC)	6316.985

Root Mean Square Error of Approximation:

RMSEA	0.028
90 Percent confidence interval - lower	0.000
90 Percent confidence interval - upper	0.052
P-value RMSEA \leq 0.05	0.936

Standardized Root Mean Square Residual:

SRMR	0.037
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)
JobSatisfaction =				
work1	1.000			
work2	1.035	0.081	12.770	0.000
work3	0.891	0.073	12.145	0.000
Wellbeing =				
happy	1.000			
mood1	1.490	0.219	6.799	0.000
mood2	0.821	0.126	6.535	0.000
DysThinking =				
perform1	1.000			
perform2	1.133	0.080	14.105	0.000
approval	0.993	0.089	11.175	0.000
ConThinking =				
beliefs	1.000			
selftalk	1.056	0.178	5.922	0.000
imagery	1.890	0.331	5.717	0.000

Regressions:

	Estimate	Std.Err	z-value	P(> z)
DysThinking -				
ConThinking	-0.131	0.078	-1.681	0.093
Wellbeing -				
DysThinking	-0.365	0.064	-5.664	0.000
ConThinking	0.067	0.061	1.096	0.273
JobSatisfaction -				
DysThinking	-0.242	0.130	-1.863	0.063
ConThinking	0.160	0.120	1.331	0.183
Wellbeing	0.797	0.202	3.946	0.000

Covariances:

	Estimate	Std.Err	z-value	P(> z)
.happy --				
.mood2	-0.043	0.018	-2.390	0.017

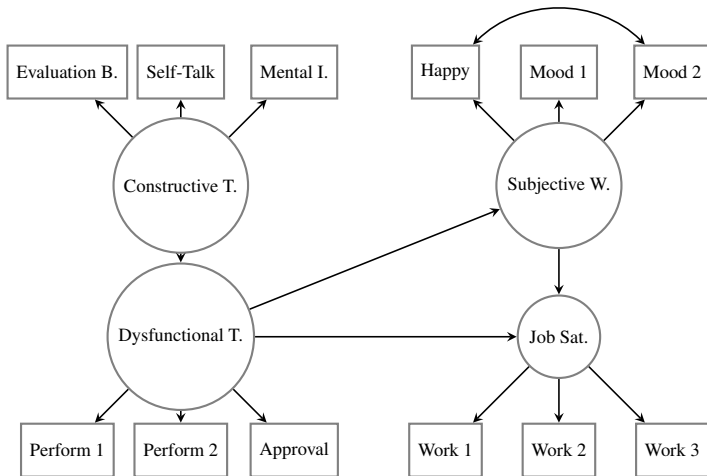
Variances:

	Estimate	Std.Err	z-value	P (> z)
.work1	0.260	0.042	6.231	0.000
.work2	0.368	0.050	7.394	0.000
.work3	0.384	0.044	8.692	0.000
.happy	0.173	0.025	6.877	0.000
.mood1	0.261	0.044	5.970	0.000
.mood2	0.178	0.022	8.133	0.000
.perform1	0.106	0.016	6.783	0.000
.perform2	0.068	0.017	3.975	0.000
.approval	0.300	0.029	10.186	0.000
.beliefs	0.292	0.043	6.862	0.000
.selftalk	1.022	0.097	10.496	0.000
.imagery	0.242	0.123	1.965	0.049
.JobSatisfaction	0.467	0.066	7.063	0.000
.Wellbeing	0.108	0.025	4.398	0.000
.DysThinking	0.231	0.031	7.579	0.000
ConThinking	0.212	0.049	4.285	0.000

- identiek dezelfde fit als voorgaand model: zelfde aantal 'paden' tussen de factoren
- SEM kan met andere woorden geen indicatie geven over de richting van de effecten!!
- we krijgen nu wel informatie over de padcoëfficiënten

- kunnen we het structureel deel nog verder vereenvoudigen?

SEM with Directed Arrows -2



Implementatie in lavaan in R:

```
semmodel2<-'JobSatisfaction =~ work1 + work2 + work3
Wellbeing      =~ happy + mood1 + mood2
DysThinking    =~ perform1 + perform2 + approval
ConThinking    =~ beliefs + selftalk + imagery
# correlated residual
happy ~~ mood2
# regressies
DysThinking    ~ ConThinking
Wellbeing      ~ DysThinking
JobSatisfaction ~ DysThinking + Wellbeing
,
```

```
fitsem2 <- cfa(semmodel2, sample.cov=COV, sample.nobs=263)
```

```
summary(fitsem2,fit.measures=T,standardized=T)
```

Output

Estimator	ML
Optimization method	NLMINB
Number of free parameters	29

Number of observations	263
------------------------	-----

Model Test User Model:

Test statistic	60.010
Degrees of freedom	49
P-value (Chi-square)	0.135

Model Test Baseline Model:

Test statistic	1087.490
Degrees of freedom	66
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.989
Tucker-Lewis Index (TLI)	0.985

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-3122.941
Loglikelihood unrestricted model (H1)	-3092.936
Akaike (AIC)	6303.882
Bayesian (BIC)	6407.474
Sample-size adjusted Bayesian (BIC)	6315.530

Root Mean Square Error of Approximation:

RMSEA	0.029
90 Percent confidence interval - lower	0.000
90 Percent confidence interval - upper	0.052
P-value RMSEA \leq 0.05	0.931

Standardized Root Mean Square Residual:

SRMR	0.043
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
JobSatisfaction =						
work1	1.000				0.787	0.840
work2	1.031	0.081	12.730	0.000	0.811	0.799
work3	0.891	0.073	12.158	0.000	0.702	0.750
Wellbeing =						
happy	1.000				0.380	0.677
mood1	1.467	0.216	6.775	0.000	0.557	0.734
mood2	0.818	0.125	6.545	0.000	0.310	0.594
DysThinking =						
perform1	1.000				0.485	0.831
perform2	1.131	0.080	14.145	0.000	0.549	0.902
approval	0.993	0.089	11.180	0.000	0.482	0.660
ConThinking =						
beliefs	1.000				0.463	0.653
selftalk	1.060	0.178	5.941	0.000	0.491	0.438
imagery	1.860	0.331	5.625	0.000	0.862	0.863

Regressions:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
DysThinking ~						
ConThinking	-0.139	0.078	-1.787	0.074	-0.133	-0.133
Wellbeing ~						
DysThinking	-0.377	0.065	-5.827	0.000	-0.482	-0.482
JobSatisfaction ~						
DysThinking	-0.258	0.131	-1.975	0.048	-0.159	-0.159
Wellbeing	0.805	0.201	3.995	0.000	0.388	0.388

Covariances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.happy --						
.mood2	-0.044	0.018	-2.457	0.014	-0.044	-0.254

Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.work1	0.259	0.042	6.175	0.000	0.259	0.295
.work2	0.372	0.050	7.454	0.000	0.372	0.361
.work3	0.382	0.044	8.651	0.000	0.382	0.437
.happy	0.171	0.025	6.716	0.000	0.171	0.542
.mood1	0.265	0.044	6.083	0.000	0.265	0.461
.mood2	0.177	0.022	8.033	0.000	0.177	0.648
.perform1	0.105	0.016	6.782	0.000	0.105	0.309
.perform2	0.069	0.017	4.042	0.000	0.069	0.186
.approval	0.300	0.029	10.185	0.000	0.300	0.564
.beliefs	0.289	0.043	6.665	0.000	0.289	0.574
.selftalk	1.017	0.097	10.441	0.000	1.017	0.808
.imagery	0.255	0.124	2.062	0.039	0.255	0.256
.JobSatisfaction	0.474	0.067	7.085	0.000	0.764	0.764
.Wellbeing	0.111	0.025	4.409	0.000	0.768	0.768
.DysThinking	0.231	0.030	7.589	0.000	0.982	0.982
ConThinking	0.215	0.050	4.262	0.000	1.000	1.000

- het effect van ‘constructive thinking’ op ‘dysfunctional thinking’ is niet significant ($p = .074$)
- ‘dysfunctional thinking’ heeft wel een significant (negatief) effect op ‘well-being’ ($p < .001$) en op ‘job satisfaction’ ($p = .048$)
- ‘well-being’ heeft een significant (positief) effect op ‘job satisfaction’ ($p < .001$)

- de hypothesen van de onderzoekers worden slechts gedeeltelijk bevestigd door de data

3.4 Modelvergelijkingen

- 2 modellen zijn genest als het ene model een submodel is van het andere
- de chi-kwadraat difference χ_D^2 laat toe de statistische significantie van de verbetering in de fit te onderzoeken als vrije parameters worden toegevoegd
- H_0 stelt dat de fit van beide modellen gelijk is
- $\chi_D^2 = \chi_s^2 - \chi_l^2$ en $df_D = df_s - df_l$ met 's' het kleiner (small) model met minder parameters en daarom meer vrijheidsgraden (df), en 'l' het groter (large) model met meer parameters en minder vrijheidsgraden
- als H_0 verworpen wordt, fit het groter model met meer vrij geschatte parameters beter dan het kleiner model dat deze parameters fixeert; als H_0 niet verworpen wordt, kunnen we kiezen voor het eenvoudiger model

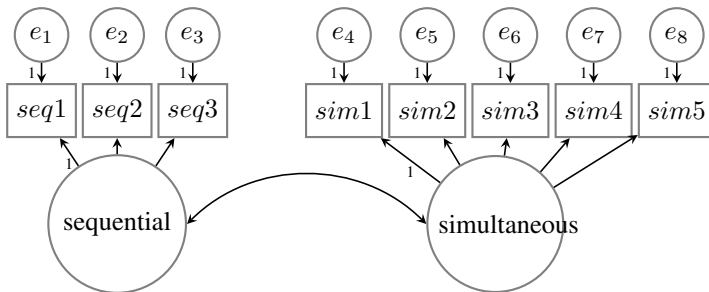
Samengevat:

Model	χ^2_M	df_M	χ^2_D	df_D	RMSEA	CFI	SRMR
<u>Measurement Model</u>							
1-factor CFA	566.80 ^a	54	-	-	.190	.498	.143
4-factor CFA	62.47 ^b	48	504.32 ^a	6	.034	.986	.040
4-factor CFA with add. res.	56.66 ^c	47	5.80 ^d	1	.028	.991	.037
<u>Structural regression model</u>							
Just-identified model (6 paths)	56.66 ^c	47	-	-	.028	.991	.037
Over-identified model (4 paths)	60.01 ^e	49	3.35 ^f	2	.029	.989	.043

^a $p < 0.001$, ^b $p = .078$, ^c $p = .158$, ^d $p = .016$, ^e $p = .135$, ^f $p = .188$

Quiz-time

Onderstaande figuur toont een CFA-model voor de 'Mental Processing Scale of the 1st Edition of Kaufman Assessment Battery for Children', een cognitieve test voor kinderen tussen 2,5 en 12,5 jaar. De ontwikkelaars van de test claimen dat de 8 subtesten voorgesteld in onderstaande figuur 2 factoren meten, 'sequential processing' en 'simultaneous processing'. De eerste 3 taken (*seq1*, . . . , *seq3*) worden verondersteld 'sequential processing' te meten, de laatste 5 taken (*sim1*, . . . , *sim5*) worden verondersteld de 'simultaneous processing' te meten.



Welk van de volgende uitspraken is correct over het voorgestelde CFA-model?

- (A) Dit is een unidimensioneel meetmodel.
- (B) Dit is een multidimensioneel meetmodel met kruisladingen maar zonder error-correlaties.
- (C) Dit is een multidimensioneel meetmodel met error-correlaties maar zonder kruisladingen.
- (D) Dit is een multidimensioneel meetmodel met error-correlaties en kruisladingen.

Referenties

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: New developments and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110.

Brenning, K., Soenens, B., Braet, C. & Bosmans, G. (2011) An adaption of the Experiences in Close Relationships Scale-Revised for use with children and adolescents. *Journal of Social and Personal Relationships*, 28, 1048-1072.

Cattell, R. B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. New York: Plenum.